

# Use of appropriate statistical tools in biomedical research: Current trend & status

Anup Kumar, Jai Kishun, Uttam Singh, Diksha Gaur, Prabhakar Mishra & C.M. Pandey

Department of Biostatistics & Health Informatics, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, Uttar Pradesh, India

Received April 21, 2020

*Background & objectives*: Due to lack of appropriate statistical knowledge, published research articles contain various errors related to the design, analysis and interpretation of results in the area of biomedical research. If research contains statistical error, however, costly, it may be of no use and the purpose of the investigation gets defeated. Many biomedical research articles published in different peer reviewed journals may retain several statistical errors and flaws in them. This study aimed to examine the trend and status of application of statistics in biomedical research articles. Study design, sample size estimation and statistical measures are crucial components of a study. These points were evaluated in published original research articles to understand the use or misuse of statistical tools.

*Methods*: Three hundred original research articles from the latest issues of selected 37 journals were reviewed. These journals were from the five internationally recognized publication groups (CLINICAL KEY, BMJ Group, WILEY, CAMBRIDGE and OXFORD) accessible through the online library of SGPGI, Lucknow, India.

*Results*: Among articles assessed under present investigation, 85.3 per cent (n=256) were observational, and 14.7 per cent (n=44) were interventional studies. In 93 per cent (n=279) of research articles, sample size estimation was not reproducible. The simple random sampling was encountered rarely in biomedical studies even though none of the articles was adjusted by design effect and, only five articles had used randomized test. The testing of assumption of normality was mentioned in only four studies before applying parametric tests.

*Interpretation & conclusions*: In order to present biomedical research results with reliable and precise estimates based on data, the role of engaging statistical experts need to be appreciated. Journals must have standard rules for reporting study design, sample size and data analysis tools. Careful attention is needed while applying any statistical procedure as, it will not only help readers to trust in the published articles, but also rely on the inferences the published articles draw.

Key words Biomedical research - design effect - sample size - statistical error - study design

The reporting of statistics in medical field becomes ever more crucial and complicated from initiation of study design to collection, management, analysis and interpretation of data and finally conclusion of the study. The increasing volume of biomedical research often leads to increasing number of contradictory findings and conclusions. Enough evidence from the literature are available to this effect<sup>1-4</sup>. Errors related to sampling technique, randomization and various other common flaws related to data analysis may lead to wrong inference<sup>5</sup>.

One of the main reasons for contradiction may occur because of sampling variability (uncertainty), as many of the studies are performed on a limited number of cases. While drawing inference through a sample, we must estimate sampling variability, *i.e.* standard error. Estimation of standard error can be done only if, the samples are drawn from the population through any probability sampling technique. Statistical data analysis of the related field is required to validate the results but, insufficient statistical knowledge perpetrates a wide range of errors related to design, execution, analysis, presentation and interpretation. Several studies have shown that statistical reporting is not appropriate<sup>6-8</sup>. The purpose of the application of statistics is not defined clearly and the researchers are not even aware of it. Every stage of a study from planning to the implementation and interpretation requires appropriate statistical knowledge. If the study is not done carefully, there may be some error associated with the application<sup>9</sup>. Different articles contributed to the production of improved statistical work by reviewing frequently found errors and flaws<sup>10-12</sup>. To answer a research question, one has to plan and execute a study to come to a valid conclusion and addressing a research question requires careful examination of various components such as, (i) study design, (ii) sample size, (iii) statistical measures, and (*iv*) inference<sup>6</sup>.

The study design for a research depends on the objective as well as the feasibility of data collection. If a researcher does not have appropriate knowledge about selection of a study design, he may get the results with low precision of estimations<sup>12</sup>. Each study design has some merits and demerits. For example, randomized controlled trials are the most powerful designs possible in medical research because, they protect from selection bias<sup>13</sup>, but their application is difficult and the results may not be generalizable. Sample size determination depends on the type of study design and some other important parameters, *i.e.* level of significance, power of the study, margin of error, and design effect related to the research question. Now, there are lots of techniques available for determining the sample size. Usually,

researchers do not investigate details of it and do this step as per previous published studies. Consequently, how they arrives at the sample size remains missing or not become clear to the readers. Once the sample size is determined, the next step is the collection of data. Data collection depends on the type of study design. Almost all the techniques of sample size determination have the assumption of obtaining data from a population by simple random sampling technique or other related variation. In biomedical research, there is a myth of random sampling. It is challenging or prohibitively expensive even to define and list the population of interest, a prerequisite for random sampling<sup>14,15</sup>. True random sampling from a population is a statistical ideal that is never attained in practice. The subjects under test in many studies are not random samples at all; instead, they are more appropriately described as samples of convenience<sup>16</sup>. Only four per cent used random sampling in an examination of 252 studies published in five biomedical journals from 1993 to 1994<sup>17</sup>. In modern day clinical and epidemiologic research, the groups of people under investigation are rarely assembled by random sampling. Hence, data collection by simple random sampling is not simple. In such case, one must adjust the sample size by design effect. Furthermore, in an interventional study design, the process of randomization at times is misinterpreted as random sampling.

There are various statistical theories in practice based on respective assumptions. The users generally apply the techniques without understanding the assumptions and hence may land into wrong conclusions<sup>18</sup>. A researcher has his/her own limitations. but one must draw conclusions based on the underlying assumptions and limitations. All the above components are interrelated to each other. Over the past few decades, use of statistics in medical research have become quite popular. The problem with inappropriate reporting of statistics is that it can have errors and flaws associated with different stages and can have an unfavourable impact on the reliability of results<sup>19</sup>. The objective of the present study was to investigate the execution of statistical methodology in biomedical research articles. Sample size estimation, study design and statistical measures were evaluated to see the current trend of using statistics in biomedical research.

# **Material & Methods**

This study analyzed articles published in 37 journals available through the e-library of Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, Uttar Pradesh, India. In total, there were eight publication house platforms in the e-library; namely, Clinical Key, BMJ Group, Cambridge, Oxford, Wiley, Informa Healthcare, Wolters Kluwer Health and Ind MED. The journals were selected randomly from a publishing group and then all original research articles from the latest available volume were downloaded and reviewed.

Sample size: It was assumed that 50 per cent (P=0.5) of the articles in medical research are likely to have an inappropriate sample size to cover the maximum sample size and to validate this two sided hypothesis, at a five per cent level of significance ( $\alpha=0.05$ ) and at the power of the study 0.80 and the absolute margin of error 10 per cent. The minimum number of articles required was 199. The formula used for sample size calculation (requirement) was to compare a one-sample proportion against a general population proportion. The result were validated using G-Power (*https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-p sychologie-und-arbeitspsychologie/gpower*) and PASS (Power Analysis and Sample Size Software) 2019 software (NCSS, LLC. Kaysville, Utah, USA).

$$n = \frac{\left(\frac{Z_{\frac{\alpha}{2}} + Z_{\beta}}{2}\right)^{2}}{\delta^{2}} (p(1-p))$$
  
$$n = \frac{\left(1.96 + 0.85\right)^{2}}{0.10^{2}} (0.5(1.85), \text{ so} \cong 198 \text{ (approx.)}$$

Considering a design effect equal to 1.5, as the selection of research articles is not exactly done by simple random sampling, required number of articles was 298.5 ( $\approx$  300).

Inclusion and exclusion criteria: Out of eight publication houses available, Clinical Key, BMJ Group, Cambridge, Oxford and Wiley were selected as rest of the others were not accessible at the time of the study. Equal allocation of 60 original articles from each publishing group, published during the specified time period from July 2018 to July 2019 were included. Short reports, letters, reviews, editorial and case studies were excluded. If original articles were not available, we skipped to other journal until the required number of articles was obtained (Figure).

According to above-mentioned criteria, 60 original articles from each of the publishing group under study were selected. In total, 300 original articles were selected and reviewed. The quality of

statistics used in articles was assessed thoroughly. Each paper was read and assessed following some norms by two experienced statisticians independently. The detailed examination of articles comprised of the explanations around sample size calculation, study design, randomization and adjustment of confounding effect of different variables and presentation of results. Usually, preciseness, robustness and efficiency of the statistical methods described in the selected original articles were assessed using yes or no assignment for every characteristic.

# Results

The outcome of the review of the articles is summarized in Tables I and II. The study designs were broadly classified under observational and interventional categories. Among 300 selected articles, 252 (84%) articles were on observational investigations, while remaining 48 (16.0%) articles presented interventional studies. Observational studies comprise of 92 (30.7%) prospective, 27 (9%) cohort based, 72 (24%) retrospective, 34 (11.3%) case-control and 27 (9%) cross-sectional studies. All interventional studies were clinical trials. Of the total reviewed articles, sample size was mentioned in 287 (95.7%) articles. Reproducibility of sample size, i.e. whether the information reported in the articles were enough to determine sample size or not, was found only in 21 (7%) articles where estimated sample sizes were supported with proper justification. On the other hand, in the remaining 266 (93%) articles, neither sample size estimation technique was mentioned precisely nor description in certain manner was provided.

Further, it was observed that, out of the total only 32 (10.6%) articles used parametric tests, whereas non-parametric tests were used in 124 (41%) articles. Forty five (15%) articles used both parametric and nonparametric tests. Ninety (30%) articles used descriptive statistics. More specifically, t test was used in 19.67 per cent, Chi-square test or Fisher exact test in 14 per cent, ANOVA in 3.66 per cent, survival analysis was used in 1.9 per cent of the articles. Only four articles (4/32; 12%) mentioned about the normality test before applying parametric test. Further, Mann-Whitney U test used in eight per cent, Wilcoxon signed-rank test in 3.34 per cent, Kruskal-Wallis H test in 3.67 per cent of the articles. Only five (2%) articles used randomized test.

# Discussion

Although among observational study designs, a cohort study allows identifying causation, in the

#### INDIAN J MED RES, APRIL 2023



Figure. Schematic diagram of selection of original articles.

Table I. Distribution of study design in selected articles			
Study design	Type of study	Frequency (%)	
Observational	Prospective/cohort	104 (34.7)/26 (8.7)	
	Retrospective/case control	75 (25)/26 (8.7)	
	Cross-sectional	25 (8.3)	
Intervention	Clinical trial	44 (14.7)	

Table II. Distribution of characteristics of the articles		
Statistical characteristics of articles	Frequency (%)	
Sample size used	287 (95.7)	
Sample size not mentioned	13 (4.3)	
Sample size not reproducible	279 (93)	
Parametric test used with normality test mentioned	4 (1.3)	
Parametric test used without normality test mentioned	28 (9.3)	
Non-parametric test used	124 (41)	
Descriptive statistics	90 (30)	
Others (Kappa, Cronbach's alpha, etc.)	3 (1)	
Both parametric and non-parametric test used	45 (15)	
Test not mentioned	7 (2.3)	
Randomized test	5 (2.0)	

reviewed articles its proportion much less was. Among observational study design, prospective study and retrospective study is done significantly more than cohort, case-control and cross-sectional study design. It appeared that prospective study design was commonly used in hospital setup as cases were available for routine follow up. The same was true for investigation following retrospective design as well. In practice, obtaining appropriate controls in observational study design seemed difficult.

Sample size was mentioned in 95 per cent of the study, but the means to estimate was not explicitly mentioned. The authors were more concerned about obtaining and reporting results, rather than going into complications of sample size determination. Once the data is collected in any study utmost care should be taken to analyse the data. Especially the study variable, its symmetry, level of measure will decide the kind of statistical analysis required to reach a conclusion. Use of various common statistical tests in biomedical research was observed in the present investigation. As most of the study was observational in nature, interpretation of the results was limited for establishing causal times. Simple random sampling is difficult to achieve in many biomedical studies as it is difficult to adopt in hospital setup. Hence, researcher should adjust the sample size by design effect. The design effect is basically the ratio of the actual variance, under the sampling method used, to the variance computed under the assumption of simple random sampling. Statistical tests developed under the assumption that the data is being collected by simple random sampling with a 100 per cent response rate, is rarely true. When untrue, one must adjust its sample size by design effect to be able to use the usual parametric and non-parametric tests. If the sample size cannot be adjusted because of cost, time and feasibility, one must use permutation or randomization test<sup>14</sup>. Randomization test is based on the sampling distribution of the test statistics obtained from all possible rearrangements of observed data set, where the critical value is the test statistics calculated from the observed data<sup>20</sup>. In this study, randomized test was found in only five articles. These tests are very uncommon as they are not available in usual statistical package.

Decision to use parametric and nonparametric tests is based on assumption about the distribution of data. Parametric tests assume a normal distribution of values, or a 'bell-shaped curve', whereas nonparametric tests are used in cases where parametric tests are not appropriate. Most nonparametric tests use some way of ranking the measurements and testing for weirdness of the distribution or where study variables are skewed/ordinal.

Based on our study findings, we suggest that before performing any statistical analysis, the researchers should be very clear about the test or methods they are using and should follow the norms and underlying assumptions so that the results obtained are significant and reliable<sup>4,21,22</sup>. Moreover, the advice of statistical experts is required before planning, designing and execution of a study. A research outcome is considered valid and can be generalized only when investigators are able to obtain a random sample of adequate size. Modern techniques are available to analyze the data in such situations. It is important to appreciate that generalizability particularly important in studies that can impact broad policy or regulatory decisions.

### Acknowledgment: None.

#### Financial support & sponsorship: None.

# Conflicts of Interest: None.

## References

- 1. Shankar PR, Varadarajan S. Statistics and ethics in medical research. *J Res Med Educ Ethics* 2018; 8:3.
- 2. Altman DG. Statistical reviewing for medical journals. *Stat Med* 1998; *17*: 2661-74.
- 3. Goodman SN. Is used in medical research. *Methods* 2002; 287: 2000-3.
- Henley SS, Golden RM, Kashner TM. Statistical modeling methods: Challenges and strategies. *Biostat Epidemiol* 2020; 4: 105-39.
- 5. Strasak AM, Zaman Q, Marinell G, Pfeiffer KP, Ulmer H. The use of statistics in medical research: A comparison of the New England journal of medicine and nature medicine. *Am Stat* 2007; *61* : 47-55.
- Ercan I, Yazici B, Yang Y, Özkaya G, Cangur S, Ediz B, et al. Misusage of statistics in medical research. Eur J Gen Med 2007; 4: 128-34.

- 7. Hanif A, Ajmal T. Statistical errors in medical journals (A critical appraisal). *Ann KEMU* 2011; *17* : 178.
- Yan F, Robert M, Li Y. Statistical methods and common problems in medical or biomedical science research. *Int J Physiol Pathophysiol Pharmacol* 2017; 9:157-63.
- Strasak AM, Zaman Q, Pfeiffer KP, Göbel G, Ulmer H. Statistical errors in medical research – A review of common pitfalls. *Swiss Med Wkly* 2007; 137: 44-9.
- 10. Glantz SA. Biostatistics: How to detect, correct and prevent errors in the medical literature. *Circulation* 1980; *61* : 1-7.
- Lee S. Avoiding negative reviewer comments: Common statistical errors in anesthesia journals. *Korean J Anesthesiol* 2016; 69: 219-26.
- Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: Critical assessment of articles in BMJ from January to March 1976. Br Med J 1977; 1: 85-7.
- Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schünemann H, *et al.* Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* 2011; 2011 : MR000012.
- Dugard P. Randomization tests: A new gold standard? J Contextual Behav Sci 2014; 3: 65-8.
- van Hoeven LR, Janssen MP, Roes KC, Koffijberg H. Aiming for a representative sample: Simulating random versus purposive strategies for hospital selection. *BMC Med Res Methodol* 2015; 15: 90.
- Geoffre K. Design and Analysis: A reseracher's handbook. 3<sup>rd</sup> ed. New Jersey: Prentice Hall PTR; 1991.
- Lundbrook J, Dudley H. Why permutation tests are superior to t and F tests in biomedical research. *The American Statistician* 1998; 52: 127-32.
- Parsons NR, Price CL, Hiskens R, Achten J, Costa ML. An evaluation of the quality of statistical design and analysis of published medical research: Results from a systematic survey of general orthopaedic journals. *BMC Med Res Methodol* 2012; *12*: 60.
- Ibrahimi E. Data analysis and documentation of statistics in biomedical research papers in Albania. *Biostat Epidemiol Int* J 2018; 1: 18-20.
- Bind MC, Rubin DB. When possible, report a fisher-exact P value and display its underlying null randomization distribution. *Proc Natl Acad Sci U S A* 2020; *117*: 19151-8.
- Laccourreye O, Fakhry N, Franco-Vidal V, Jankowski R, Karkas A, Leboulanger N, *et al.* Statistics in scientific articles published in the European Annals of Otorhinolaryngology Head & Neck Diseases. *Eur Ann Otorhinolaryngol Head Neck Dis* 2020; *138*: 89-92.
- 22. Indrayan A. Reporting of basic statistical methods in biomedical journals: Improved SAMPL guidelines. *Indian Pediatr* 2020; *57* : 43-8.

For correspondence: Dr Jai Kishun, Department of Biostatistics & Health Informatics, 5<sup>th</sup> Floor, Central Library Complex, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Raebareli Road, Lucknow 226 014, Uttar Pradesh, India e-mail: jaikishan.stat@gmail.com